

## Claims

What is claimed is:

1. A process of identifying transcription factors such as activators and/or repressors comprising:  
providing cells with a nucleic acid sequence at least comprising a sequence CACCT (the first 5 nucleotides of SEQ ID NO: 1), preferably twice a CACCT sequence (the first 5 nucleotides of SEQ ID NO: 1), as bait(s) for the screening of a library encoding potential transcription factors and  
performing a specificity test to isolate said transcription factors.
2. A process of identifying transcription factors such as activators and/or repressors comprising:  
providing cells with a nucleic acid sequence comprising one of the sequences CACCT-N-CACCT (SEQ ID NO: 1), CACCT-N-AGGTG (SEQ ID NO: 2), AGGTG-N-CACCT (SEQ ID NO: 3), or AGGTG-N-AGGTG (SEQ ID NO: 4) as bait wherein N is a spacer sequence.
3. A process according to claim 1 or claim 2 wherein the transcription factor comprises separated clusters of zinc fingers.
4. A process according to claim 1, claim 2, or claim 3 wherein the sequence originates from a promoter region.
5. A process according to claim 4 wherein the promoter region is selected from the group consisting of Brachyury,  $\alpha$ 4-integrin, follistatin, and E-cadherin.
6. A transcription factor produced by the process of claim 1, claim 2, claim 3, claim 4, or claim 5.

7. A process for identifying compounds with an interference capability towards transcription factors as defined in claim 6 by adding a sample comprising a potential compound to be identified to a test system comprising:
- (i) a nucleotide sequence comprising one of the sequences CACCT-N-CACCT (SEQ ID NO: 1), CACCT-N-AGGTG (SEQ ID NO: 2), AGGTG-N-CACCT (SEQ ID NO: 3), or AGGTG-N-AGGTG (SEQ ID NO: 4) as bait wherein N is a spacer, and (ii) a protein capable to bind said nucleotide sequence,
- incubating said sample in said system for a period of time sufficient to permit interaction of the compound or its derivative or counterpart thereof with said protein,
- comparing the amount and/or activity of the protein bound to the nucleotide sequence before and after said adding and
- identification and optionally isolation and/or purification of the compound.
8. The process according to claim 7 wherein the protein is a Smad-interacting protein.
9. The process according to claim 8, wherein said Smad-interacting protein is SIP1.
10. A compound produced by the process of claim 7, claim 8, or claim 9.
11. The compound of claim 10, wherein said compound modifies regulation of E-cadherin expression by SIP1.
12. A pharmaceutical composition to prevent tumor invasion and/or metastasis, said pharmaceutical composition comprising:
- the compound of claim 10 or claim 11 in an amount to prevent tumor invasion and/or metastasis in a subject, and
- a pharmaceutically acceptable excipient.
13. A test kit to perform the process of claim 7, said test kit comprising:

a nucleotide sequence comprising a sequence selected from the group consisting of CACCT-N-CACCT (SEQ ID NO: 1), CACCT-N-AGGTG (SEQ ID NO: 2), AGGTG-N-CACCT (SEQ ID NO: 3), and AGGTG-N-AGGTG (SEQ ID NO: 4) wherein N is a spacer sequence and

(ii) a protein capable of binding said nucleotide sequence.

14. A test kit to perform the process of claim 2, said test kit comprising:

a nucleic acid sequence comprising one of the sequences CACCT-N-CACCT (SEQ ID NO: 1), CACCT-N-AGGTG (SEQ ID NO: 2), AGGTG-N-CACCT (SEQ ID NO: 3), or AGGTG-N-AGGTG (SEQ ID NO: 4), wherein N is a spacer sequence.

15. A method for detecting an interaction between a first interacting protein and a second interacting protein comprising:

providing a suitable host cell with a first fusion protein comprising a first interacting protein fused to a DNA binding domain capable to bind a nucleic acid sequence comprising one of the sequences CACCT-N-CACCT (SEQ ID NO: 1), CACCT-N-AGGTG (SEQ ID NO: 2), AGGTG-N-CACCT (SEQ ID NO: 3), or AGGTG-N-AGGTG (SEQ ID NO: 4) wherein N is a spacer sequence,

providing said suitable host cell with a second fusion protein comprising a second interacting protein fused to a DNA binding domain capable to bind a nucleic acid sequence comprising one of the sequences CACCT-N-CACCT (SEQ ID NO: 1), CACCT-N-AGGTG (SEQ ID NO: 2), AGGTG-N-CACCT (SEQ ID NO: 3) or AGGTG-N-AGGTG (SEQ ID NO: 4) wherein N is a spacer sequence,

subjecting said host cell to conditions under which the first interacting protein and the second interacting protein are brought into close proximity and determining whether a detectable gene present in the host cell and located adjacent to said nucleic acid sequence has been expressed to a greater degree than if expressed in the absence of the interaction between the first and the second interacting protein.

16. An isolated nucleic acid sequence comprising a sequence selected from the group consisting of CACCT-N-CACCT (SEQ ID NO: 1), CACCT-N-AGGTG (SEQ ID NO: 2), AGGTG-N-CACCT (SEQ ID NO: 3), and AGGTG-N-AGGTG (SEQ ID NO: 4) wherein N is a spacer.

17. A method of identifying a new target gene, said method comprising:  
identifying said new target gene using a nucleic acid sequence, said nucleic acid sequence comprising a sequence selected from the group consisting of CACCT (the first five nucleotides of SEQ ID NO: 1), CACCT-N-CACCT (SEQ ID NO: 1), CACCT-N-AGGTG (SEQ ID NO: 2), AGGTG-N-CACCT (SEQ ID NO: 3), and AGGTG-N-AGGTG (SEQ ID NO: 4) wherein N is a spacer.